

Modelos 3D de reconstrucción facial: Una breve revisión

Victor Hernández-Manrique¹, Miguel González-Mendoza¹,
Carlos Vilchis¹, Mauricio Méndez-Ruiz²,
Carmina Pérez-Guerrero²

¹ Tecnológico de Monterrey,
Escuela de Ingeniería y Ciencias,
México

² Eugenia Virtual Humans S.A. de C.V.,
Laboratorio de Investigación,
México

{A01731594, carlos.vilchis, mgonza}@tec.mx,
{mauricio, carmina}@eugenia.tech

Resumen. Los algoritmos de reconstrucción facial 3D han demostrado ser una herramienta poderosa para diversas aplicaciones, incluido el reconocimiento facial, la realidad virtual y las imágenes médicas. Sin embargo, la complejidad y los recursos computacionales requeridos para estos algoritmos, así como la restringida disponibilidad de los conjuntos de datos (*datasets*), han limitado su accesibilidad a investigadores y profesionistas. Esto es desafortunado porque la reconstrucción facial en 3D tiene el potencial de beneficiar a una gama más amplia de individuos y comunidades, incluidos aquellos que forman parte de sectores como la atención médica, entretenimiento y seguridad. La siguiente investigación indaga en la evolución de esta tecnología, así como en su accesibilidad para las masas. Se presenta una tabla comparativa de tres modelos 3D de reconstrucción facial con el fin de proporcionar las características básicas de cada método, lo que permite a cualquiera seleccionar un algoritmo que se adapte a las propiedades deseadas a través de una comparación directa de los atributos clave.

Palabras clave: Reconstrucción facial, modelos 3D, MICA, 3DDFA-V2, SynergyNet, dimensiones algorítmicas, generación de avatares.

A Brief Review on 3D Models for Facial Reconstruction

Abstract. 3D face reconstruction algorithms have proven to be a powerful tool for various applications, including facial recognition, virtual reality, and medical imaging. However, the complexity and computational resources required for these algorithms, as well as the restricted availability of datasets, have limited their accessibility to a select group of researchers and professionals. This is unfortunate because 3D face reconstruction has the potential to benefit a broader range of individuals and communities, including those in healthcare,

entertainment, and security. The following research dives into the evolution of this technology, as well as its accessibility to the masses. A comparative table of three 3D face reconstruction models is presented in order to provide the base characteristics of each method, allowing anyone to select an algorithm that suits the features they are looking for through a straightforward comparison of key attributes.

Keywords: 3D face reconstruction, MICA, 3DDFA-V2, SynergyNet, algorithmic dimensions, avatar generation.

1. Introducción

Durante el último siglo, la reconstrucción facial en 3D se ha convertido en un campo de investigación en rápida evolución, con el objetivo de crear modelos 3D de humanos realistas y precisos [7]. Este desarrollo no sería posible sin avances en áreas como la visión computacional o el aprendizaje automático, impulsados por su diversa gama de aplicaciones.

El sector del entretenimiento ofrece un nuevo y fascinante uso para la reconstrucción facial en 3D [6]. La capacidad de producir modelos 3D realistas de rostros humanos es cada vez más crucial para el desarrollo de tecnologías como la realidad virtual y aumentada. Esta tecnología permite el mapeo en tiempo real del propio rostro del usuario en un avatar digital, la cual sirve para experiencias virtuales inmersivas como videojuegos y conciertos de realidad virtual [12].

Además, la industria cinematográfica puede beneficiarse de la reconstrucción facial en 3D para duplicar a los artistas digitalmente, ya que podría ser usado para acrobacias, filmar situaciones que serían demasiado arriesgadas o muy costosas de realizar.

La atención médica es otra industria donde la reconstrucción facial en 3D se ve muy prometedora [13]. Los médicos e investigadores pueden examinar una variedad de deformidades faciales y problemas médicos con esta tecnología. Por ejemplo, los procedimientos difíciles como la cirugía de labio leporino y paladar hundido se pueden planificar y practicar utilizando modelos 3D de anatomía facial.

En los últimos años, las técnicas de aprendizaje profundo, siendo las redes neuronales convolucionales (*CNN*, por sus siglas en inglés) uno de los conceptos más revolucionadores en este campo, presentaron nuevos desafíos en este tema. El estudio y desarrollo de modelos ligeros para la creación de rostros humanos en 3D con bajo costo computacional ha llegado a llamar la atención de todos. La investigación en esta área digital abriría nuevas vías de conexión, teniendo como punto de partida el Metaverso [17].

2. Planteamiento del problema

Muchas dificultades persisten a pesar de los avances realizados en el campo de la reconstrucción facial 3D [1]. La ausencia de datos de entrenamiento es uno de los principales problemas.

Para lograr aprender las diferencias en la geometría facial y la textura, los sistemas de reconstrucción facial 3D se basan principalmente en los datos de entrenamiento. El proceso de obtener escaneos faciales 3D precisos a partir de un conjunto de datos extensos y variados sigue siendo desafiante y requiere mucho tiempo.

La precisión de los modelos reconstruidos puede verse afectada por el sobreajuste y la mala generalización debido a esta escasez de datos.

La complejidad computacional de los algoritmos es otra dificultad [20]. Las técnicas modernas a veces necesitan mucha potencia de procesamiento, lo que las hace inadecuadas para aplicaciones en tiempo real o dispositivos con capacidad de procesamiento limitada.

El requerimiento de métodos avanzados de registro y optimización, así como el manejo de estructuras de datos grandes y complicadas, lleva a esta complejidad computacional. Por lo tanto, todavía es difícil crear algoritmos ligeros y efectivos que puedan producir modelos 3D precisos en tiempo real.

La robustez a los cambios en la iluminación y las expresiones faciales es un requisito indispensable [14]. Incluso al capturar rostros con expresiones faciales extremas, los algoritmos de reconstrucción facial 3D deben ser capaces de manejar una variedad de situaciones de iluminación y registrar con precisión la geometría y textura facial. Sin embargo, la precisión de los modelos reconstruidos puede verse considerablemente afectada por estos cambios. Por lo tanto, para que la reconstrucción tenga éxito en aplicaciones prácticas, los enfoques que pueden manejar estas variaciones de manera robusta son esenciales.

Asimismo, la reconstrucción facial en 3D se ve desafiada por consideraciones de ética y privacidad [16]. Las preocupaciones sobre la seguridad y la privacidad pueden surgir cuando es posible producir representaciones 3D de alta calidad de las caras de las personas, especialmente cuando se utiliza un software de realidad virtual o reconocimiento facial. Dicho uso indebido de la tecnología puede resultar en posibles violaciones de privacidad, robo de identidad u otros fallos de seguridad.

3. Trabajo relacionado

3.1. Dimensiones algorítmicas

Los parámetros específicos, configuraciones y las decisiones de diseño que tienen un impacto en el comportamiento y la efectividad de un algoritmo se conocen como dimensiones algorítmicas [8]. La selección de la técnica de optimización, el tamaño y la estructura de la red neuronal, la selección de la función de pérdida, el número de capas y las funciones de activación utilizadas en cada capa son algunos ejemplos de estas dimensiones.

Los investigadores pueden afinar el comportamiento del algoritmo para ajustarse mejor al dominio del problema en particular y mejorar su rendimiento cambiando estas dimensiones. En general, un algoritmo de reconstrucción facial 3D se compone de un modelo de detección de rostros que marca el área donde se encuentra un rostro en la imagen de entrada. Los puntos de referencia faciales se obtienen para señalar objetos como labios, cejas, ojos, etc.

La etapa de alineación permite que el algoritmo tenga una base de referencia para el enmascaramiento. Se extraen las características de apariencia y una reducción de dimensionalidad obtiene los elementos más importantes para la escultura de un rostro humano en un software [15].

La precisión, la velocidad y el costo de computación de un modelo, características comunes de estos métodos, pueden verse fuertemente afectados por las dimensiones algorítmicas que se utilizan. Por ejemplo, una optimización en el costo computacional permitiría la operación de un modelo desde un hardware especializado a uno común.

3.2. Modelos

El *NoW Challenge* recopiló métodos de reconstrucción facial 3D de última generación basados en un conjunto de datos de referencia de escaneos faciales 3D e imágenes 2D correspondientes con diversos grados de dificultad [10].

El objetivo era establecer una medida de rendimiento consistente para la precisión y robustez de los modelos. Las evaluaciones de esta competencia consistieron en un rendimiento no métrico, donde se tuvo en cuenta la suavidad de la superficie o el nivel de detalle. El rendimiento métrico proporcionó un conjunto de resultados numéricos en los que los métodos se compararon entre sí.

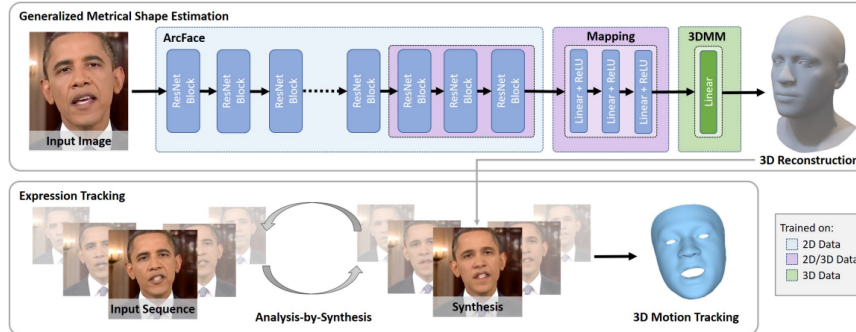
MICA. El modelo *Metric Face*, también conocido como MICA, es un método propuesto por Zielonka et al. en el 2022 [19]. Utilizaron una red de reconocimiento facial previamente entrenada en un conjunto de datos de imágenes 2D a gran escala para entrenar su estimador de forma facial de manera supervisada, lo que resulta en una mejor precisión que los métodos actuales de última generación.

A pesar de que este método tiene los mejores resultados en el NoW Challenge, no proporciona los puntos de referencia o la textura UV de una imagen de entrada. Su arquitectura, mostrada en la figura 1a, utiliza una versión modificada de ArcFace, un modelo de reconocimiento facial, como su codificador de identidad. Se realiza un mapeo 2D/3D después del proceso de ArcFace, entrenando datos 3D en un modelo lineal para su aplicación en los modelos transformables en 3D (3DMM, por sus siglas en inglés).

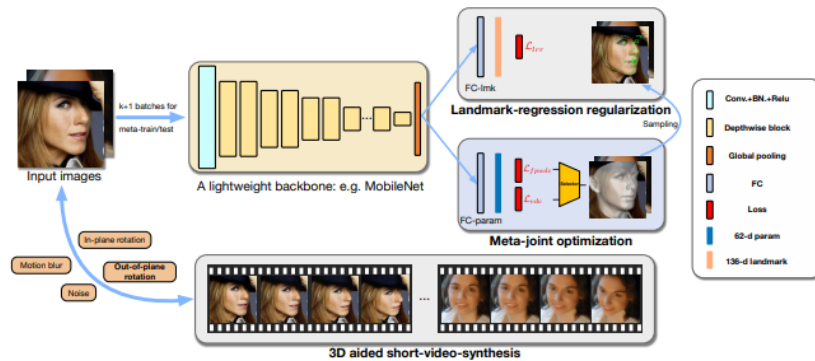
3DDFA-V2. La segunda versión de *3D Dense Face Alignment* (Alineación Densa de Rostros 3D), también conocida como 3DDFA-V2, equilibra la velocidad, la precisión y la estabilidad en comparación con su predecesora [3]. Utiliza MobileNet como un backbone ligero para predecir parámetros 3DMM (vistos en la figura 1b).

La regularización de regresión de puntos de referencia (*Landmark Regression Regularization* - LRR) proporciona parámetros que no se consideraron para regresar los puntos de referencia. La optimización del meta-conjunto (*Meta-joint Optimization*) combina el paso LRR con el rápido costo de distancia ponderado del parámetro (*fast Weighted Parameter Distance Cost* - fWPDC) y el costo de distancia de vértice (*Vertex Distance Cost* - VDC), el cual selecciona y actualiza los estados de los parámetros dentro del modelo.

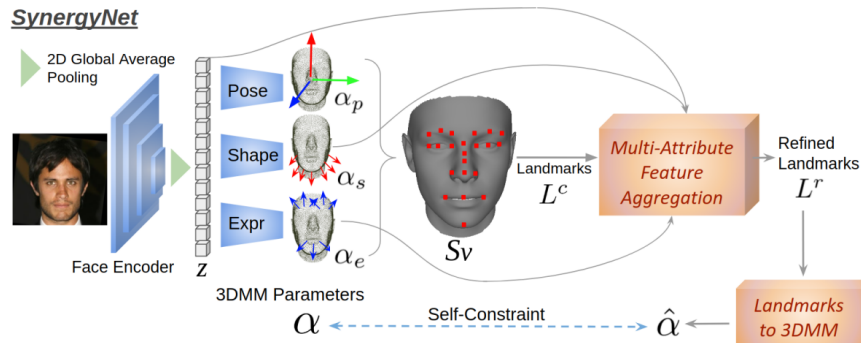
Finalmente, el análisis de video corto asistido por 3D (*3D aided short-video analysis*) simplemente expande una imagen para formar varios fragmentos.



(a) Arquitectura de MICA [19].



(b) Arquitectura de 3DDFA-V2 [3].



(c) Arquitectura de SynergyNet [18].

Fig. 1. Modelos 3D de reconstrucción facial.

SynergyNet. SynergyNet, desarrollado por Wu et al. en el 2021 [18], propone un proceso de sinergia entre los modelos transformables 3D (3DMM) y los puntos de referencia faciales 3D para predecir la geometría, la textura UV y, lógicamente, los puntos de referencia de cualquier entrada.

Tabla 1. Comparativa de dimensiones algorítmicas entre métodos.

Dimensiones Algorítmicas	MICA	3DDFA-V2	SynergyNet
Método de construcción	Reconstrucción métrica	Regresión Profunda + Refinamiento de Malla	Combinación entre 3DMM y puntos de referencia 3D
Complejidad del modelo	Alta	Baja	Moderado
Complejidad computacional	Alta	Baja - Media	Media - Alta
Robustez	Sensible a la orientación de la cara y la iluminación	Sensible a la orientación de la cara y la iluminación	Estable frente al cambio en la orientación de la cara y la iluminación
Precisión	Alto	Alto	Alto
Rendimiento	Rápido	En tiempo real	Rápido

Los puntos de referencia 3D se extraen y son refinados a partir de mallas faciales que se construyeron con parámetros 3DMM, las cuales son pose, forma y expresión. En comparación con 3DDFA-V2, el backbone en esta arquitectura, mostrado en la figura 1c, usa MobileNetV2.

3.3. Comparación de dimensiones algorítmicas entre los modelos 3D de reconstrucción facial

En el *NoW Challenge* [10], MICA obtuvo los mejores resultados tanto en la evaluación métrica como no métrica, con una mediana (en milímetros) de 1,08 y 0,90, respectivamente. 3DDFA-V2 y SynergyNet mantuvieron sus posiciones, siendo la última mencionada con los peores resultados en comparación con las otras dos.

Vale la pena mencionar que estos métodos tienen los procesos de instalación más sencillos en comparación con los otros algoritmos que se encuentran en el *NoW Challenge*. La sencillez de instalación y la compatibilidad de sus librerías fueron la razón principal por la que se eligieron estos algoritmos. Otros modelos no eran posibles de ejecutar ya que sus dependencias eran obsoletas o las versiones de librerías entraban en conflicto.

4. Discusión

Los aspectos algorítmicos de las tres propuestas demuestran que cada método tiene distintas ventajas y desventajas, como fue visto en la tabla 1. Debido a su complejidad de modelo relativamente alta, MICA requiere mucha potencia computacional para ejecutarse. Sin embargo, para lograr su gran precisión en la reconstrucción de geometría facial 3D, se requiere una complejidad considerable.

Por el contrario, el modelo 3DDFA-V2 es ligero y tiene un bajo nivel de complejidad, lo que permite velocidades de procesamiento más altas, pero posiblemente comprometiendo la precisión del resultado final.

SynergyNet logra una mezcla sólida entre las otras dos técnicas gracias a su moderada complejidad del modelo y gran precisión en la reconstrucción de la geometría facial 3D. Las principales diferencias de cada método se encuentran en sus limitaciones. MICA no presenta un análisis de los puntos de referencia o la textura UV de una imagen de entrada.

3DDFA-V2 lo hace, pero falla al no predecir la forma total de la cabeza correctamente. Finalmente, SynergyNet no puede producir una reconstrucción facial 3D basada en una imagen propia. Sus resultados son mallas 3D de datos internos. El trabajo a futuro se basará en el estudio de modelos ligeros que podrían reemplazar el backbone actual de alguno de los métodos para reducir el costo computacional sin comprometer la precisión del algoritmo.

Los modelos fueron entrenados y probados en diversos procesadores NVIDIA, por lo que se requiere un GPU de esta marca para ejecutar los algoritmos. Esto es importante de destacar y plantea la necesidad de reducir el costo computacional de los modelos, ya que no todos tienen la posibilidad de comprar este tipo de GPUs exigentes.

Brindar acceso a herramientas como los algoritmos de reconstrucción facial en 3D es esencial para garantizar que todos tengan la misma oportunidad de participar en diversas actividades que requieren dicha tecnología. Un área donde la reconstrucción facial en 3D se ha vuelto cada vez más relevante es en el campo de la medicina.

En procedimientos médicos como la cirugía de reconstrucción facial o la ortodoncia, la reconstrucción facial 3D puede proporcionar información valiosa y ayudar a los médicos a tomar decisiones vitales [13]. Brindar acceso a herramientas de reconstrucción facial 3D a todos puede ayudar a las personas de diferentes orígenes a recibir un mejor tratamiento, independientemente de su estatus financiero o social.

Utilizando la reconstrucción facial 3D en la educación, se pueden crear módulos de aprendizaje interactivos que pueden ayudar a los estudiantes a comprender mejor temas difíciles [2]. Por último, la reconstrucción facial en 3D en la aplicación de la ley puede ayudar con las investigaciones forenses y acelerar el proceso de búsqueda de personas desaparecidas o sospechosas [9].

Pero para que las técnicas de reconstrucción facial en 3D sean accesibles para todos, es necesario resolver una serie de problemas, incluido el alto costo computacional y la escasez de datos de entrenamiento. El desarrollo de algoritmos más precisos y eficientes que puedan ejecutarse en muchas plataformas de hardware y software debe seguir siendo el objetivo de los investigadores y desarrolladores. Adicionalmente, se deben hacer esfuerzos para aumentar la facilidad de uso de estas tecnologías para las personas que podrían carecer de los medios o la experiencia para hacerlo.

5. Conclusión

Esta investigación presentó una comparación de dimensiones algorítmicas de tres modelos de reconstrucción facial 3D de última generación. Su principal similitud se encuentra en la precisión de estos métodos, pero los diversos procesos que realiza cada algoritmo o sus complejidades son características importantes que un usuario debe tener en cuenta al momento de elegir el procedimiento que mejor se ajuste a sus necesidades, así como las limitaciones que se discutieron.

Como se mencionó anteriormente, la disponibilidad de recursos informáticos y técnicas de aprendizaje automático ha llevado a un gran avance en los algoritmos de reconstrucción facial 3D en los últimos años.

Pero también ha habido más dificultades en la creación de modelos 3D precisos y realistas de rostros humanos a un bajo costo computacional [5]. La ausencia de conjuntos de datos extensos y diversos, la complejidad de las expresiones faciales y la necesidad de lograr un equilibrio entre precisión y eficiencia computacional son solo algunas de estas dificultades.

Por lo tanto, la creación de modelos ligeros para la reconstrucción de rostros en 3D es crucial para superar estas dificultades. Al lograr esto, sería posible poner estos modelos en uso en hardware estándar, como teléfonos inteligentes o tabletas, abriendo la tecnología a un público más amplio. Esta accesibilidad sería ventajosa en una variedad de industrias donde la reconstrucción facial 3D tiene la capacidad de aumentar y actualizar las aplicaciones actuales [4, 11].

Garantizar que las personas de diversos orígenes y estatus socioeconómicos puedan beneficiarse de los avances de estas aplicaciones por igual requiere otorgar acceso a estas herramientas y tecnología a todos.

Fomentar la alfabetización digital y proporcionar a los usuarios las herramientas que necesitan para generar y utilizar adecuadamente los modelos de reconstrucción facial en 3D implica poner a disposición el hardware y el software necesarios, así como proporcionar recursos de capacitación e instrucción.

La creación de modelos portátiles de reconstrucción facial 3D y la expansión de su comerciabilidad pueden resultar en avances sustanciales en una serie de industrias y dar a todos la misma oportunidad de beneficiarse de estas tecnologías de vanguardia.

Agradecimientos. Los autores quieren agradecer el apoyo financiero del Tecnológico de Monterrey a través del programa “Challenge-Based Research Funding Program 2022”. Project ID # E120 - EIC-GI06 - B-T3 - D.

Referencias

1. Feng, M., Gilani, S. Z., Wang, Y., Mian, A.: 3D face reconstruction from light field images: A model-free approach. In: Proceedings of the European Conference on Computer Vision (ECCV), vol. 11214, pp. 501–518 (2018) doi: 10.1007/978-3-030-01249-6_31
2. Goeser, P. T., Hamza-Lup, F. G., Johnson, W. M., Scharfer, D.: VIEW: A virtual interactive web-based learning environment for engineering. *Advances in Engineering Education*, vol. 2, no. 3, pp. 1–24 (2011) doi: 10.48550/arXiv.1811.07463
3. Guo, J., Zhu, X., Yang, Y., Yang, F., Lei, Z., Li, S. Z.: Towards fast, accurate and stable 3D dense face alignment. In: Computer Vision- European Conference on Computer Vision (ECCV 2020), pp. 152–168 (2020) doi: 10.1007/978-3-030-58529-7_10
4. Jin, H., Wang, X., Zhong, Z., Hua, J.: Robust 3D face modeling and reconstruction from frontal and side images. *Computer Aided Geometric Design*, vol. 50, pp. 1–13 (2017) doi: 10.1016/j.cagd.2016.11.001
5. Korban, M., Li, X.: A survey on applications of digital human avatars toward virtual co-presence (2022)

6. Molina, L.: Celebrity avatars: A technical approach to creating digital avatars for social marketing strategies. Ph. D. thesis, Florida Atlantic University (2021)
7. Morales, A., Piella, G., Sukno, F. M.: Survey on 3D face reconstruction from uncalibrated images. *Computer Science Review*, vol. 40 (2021) doi: 10.1016/j.cosrev.2021.100400
8. Ngu, A.: Dimensional complexity and algorithmic efficiency. *International Journal of Modern Nonlinear Theory and Application*, vol. 11, no. 1 (2021) doi: 10.4236/ijmnta.2022.111001
9. Raneri, D.: Enhancing forensic investigation through the use of modern three-dimensional (3D) imaging technologies for crime scene reconstruction. *Australian Journal of Forensic Sciences*, vol. 50, no. 6, pp. 697–707 (2018) doi: 10.1080/00450618.2018.1424245
10. Sanyal, S., Bolkart, T., Feng, H., Black, M.: Learning to regress 3D face shape and expression from an image without 3D supervision. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7763–7772 (2019) doi: 10.1109/CVPR.2019.00795
11. Sevastopolsky, A., Ignatiev, S., Ferrer, G., Burnaev, E., Lempitsky, V.: Relightable 3D head portraits from a smartphone video (2020)
12. Slater, M., Cabrera, C., Senel, G., Banakou, D., Beacco, A., Oliva, R., Gallego, J.: The sentiment of a virtual rock concert. *Virtual Reality*, pp. 1–25 (2022) doi: 10.1007/s10055-022-00685-9
13. Stern, G., Fu, Z., Ardabilian, M.: 3D face analysis for healthcare. *Biometrics under Biomedical Considerations*, pp. 147–160 (2019) doi: 10.1007/978-981-13-1144-4_6
14. Tran, A. T., Hassner, T., Masi, I., Paz, E., Nirkin, Y., Medioni, G.: Extreme 3D face reconstruction: Seeing through occlusions. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3935–3944 (2018) doi: 10.1109/CVPR.2018.00414
15. Vilchis, C., Perez-Guerrero, C., Mendez-Ruiz, M., Gonzalez-Mendoza, M.: A survey on the pipeline evolution of facial capture and tracking for digital humans. *Multimedia Systems*, pp. 1–24 (2023) doi: 10.1007/s00530-023-01081-2
16. Vladimirov, I., Nenova, M., Nikolova, D., Terneva, Z.: Security and privacy protection obstacles with 3D reconstructed models of people in applications and the metaverse: A survey. In: *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*, pp. 1–4 (2022) doi: 10.1109/ICEST55168.2022.9828791
17. Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T. H., Shen, X.: A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 319–352 (2022) doi: 10.1109/COMST.2022.3202047
18. Wu, C. Y., Xu, Q., Neumann, U.: Synergy between 3DMM and 3D landmarks for accurate 3D facial geometry. In: *2021 International Conference on 3D Vision (3DV)*, pp. 453–463 (2021) doi: 10.1109/3DV53792.2021.00055
19. Zielonka, W., Bolkart, T., Thies, J.: Towards metrical reconstruction of human faces. In: *Computer Vision- European Conference on Computer Vision*, vol. 13673, pp. 250–269 (2022) doi: 10.1007/978-3-031-19778-9_15
20. Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., Stamminger, M., Nießner, M., Theobalt, C.: State of the art on monocular 3D face reconstruction, tracking, and applications. In: *Computer Graphics Forum*, vol. 37, pp. 523–550 (2018) doi: 10.1111/cgf.13382